Chapter 10

Discussion

10.1 Discussion

The aim of the research presented in this thesis was to contribute to the scientific knowledge underlying the prediction of academic achievement in college admissions, given the current practical and legal constraints in the Netherlands and to contribute to effective college admission procedures in general. As discussed in chapter 1, effective admission procedures should at least meet the following requirements: (1) they should have good predictive validity for the outcomes of interest; (2) they should be fair, that is, they should be unbiased against gender, ethnicity, and SES, and (3) they should be perceived as favorable and fair by stakeholders. Throughout this thesis, academic achievement was defined as the main outcome of interest. Admittedly, there are many other possible predictors for effective admission procedures and also many other possible outcomes of interest. However, I think that there is a general agreement that the requirements described above are important for effective admission procedures and that academic achievement is an important outcome variable. This chapter provides a discussion of the findings presented in this thesis. I first discuss the findings in light of the different predictors that are often used or suggested to be used in college admissions. Second, I reflect on the effects of selective admission in the Netherlands and discuss topics for future research.

10.1.1 Previous Educational Achievement

The results presented in chapters 2 and 3 confirmed that high school GPA was a good predictor of short- and long-term GPA, and to a lesser extent, of academic progress and retention (e.g., Westrick et al., 2015). Differential prediction and bias of high school GPA was not studied in this thesis, but previous research found that high school grades showed some underprediction of female academic performance and some overprediction of academic performance for ethnic minority students, but to a lesser extent than cognitive ability test scores (e.g., Mattern et al., 2008; Zwick, 2017). Surprisingly, the findings described in chapter 6 showed that high school GPA was perceived unfavorably by applicants for use in admission and matching procedures and was rated low on *study-relatedness*, *chance to perform*, *applicant differentiation*, and *face validity,* indicating that applicants do not view their high school grades as relevant when judging their fit to a specific academic program. In addition, the use of high school grades in admission procedures is hindered by a lack of comparability across schools and countries and the Dutch legal restrictions that prohibit using grades as the only admission criterion.

10.1.2 Noncognitive Measures

Noncognitive characteristics such as personality traits and motivation are increasingly popular in admission procedures; in the Netherlands, about 70% of

the students who participated in a matching procedure and over half of the students who participated in selective admission indicated that such skills were assessed in the admission procedure (van den Broek et al., 2017; Warps et al., 2017). These characteristics are usually measured with self-report questionnaires, and are mostly assessed *in addition to* more cognitively-oriented abilities and skills tests. As discussed in chapter 9, there are three main reasons to include such measures in admission procedures: (1) noncognitive measures predict important outcomes other than academic exam performance (Lievens, 2013; Schmitt, 2012); (2) noncognitive measures have incremental validity over cognitive measures (Richardson et al., 2012; Robbins et al., 2004), and (3) noncognitive measures reduce adverse impact and differential prediction (Keiser et al., 2016; Mattern, Sanchez, & Ndum, 2017).

Indeed, there is evidence in the literature that including noncognitive measures can have benefits. However, as discussed in several chapters in this thesis, we should be careful to incorporate these characteristics in high-stakes admission procedures. As demonstrated in chapter 8, using an SJT measuring interpersonal skills to predict outcomes other than academic achievement in medical school may show little incremental utility. One of the reasons that noncognitive measure do not seem to fulfill their promise of reduced adverse impact and increased validity in operational selection settings (e.g., MacKenzie, Dowell, Ayansina, & Cleland, 2017; Morgeson et al., 2007a; Zwick, 2017) is the opportunity to engage in impression management or faking (Birkeland et al., 2006). The self-report format used to measure noncognitive constructs is the Achilles heel of noncognitive assessment. Several authors have argued that impression management is a non-issue that does not affect validity (Ones et al., 1996, 2007). In contrast, the results in chapter 5 showed that the predictive- and incremental validity of self-reported noncognitive traits and skills were attenuated when they were obtained in a high-stakes admission context, most likely due to impression management and faking. As shown in Chapter 6, the possibility to fake does not only affect the predictive validity of such instruments, but also their perceived favorability by applicants; personality- and motivation questionnaires were rated modestly favorably, but were perceived less favorably when they were judged as easy to fake.

In the literature there are many studies of methods to overcome impression management in noncognitive assessment. Studies on the use of forced-choice items are currently the most popular; several authors claimed that forced-choice items are 'fake-proof' (e.g., Hirsh & Peterson, 2008; Stark, Chernyshenko, & Drasgow, 2011). Chapters 5 and 9 provided several arguments why that claim is at best premature. Here, I briefly mention the main drawbacks of the forced-choice format

that hinder successful large-scale implementation in high-stakes operational settings. Forced-choice questionnaires are more difficult to construct and to score than Likert-scale questionnaires, and thus require more resources and financial support; they have a higher cognitive load that could potentially affect their incremental validity over cognitive measures (e.g., Christiansen et al., 2005); and their operational use in high-stakes assessment would require large item pools to prevent the items from 'getting out' (Kyllonen, 2017). Especially this last requirement is difficult to realize for the relatively narrow noncognitive constructs that are often assessed with these measures. So, at the moment, I conclude that noncognitive traits and skills may show predictive- and incremental validity to academic achievement, but it is difficult to measure them validly – as separate entities– in high-stakes admission procedures. Contrary to some claims, there is no solution to this problem yet.

10.1.3. Curriculum sampling

Curriculum-sampling tests are a relatively novel development in admission testing, and they are becoming increasingly popular in European admission procedures. The results in chapters 2 and 3 showed that curriculum-sampling tests were good predictors of academic achievement in multiple cohorts of applicants to a psychology program. A curriculum-sampling test consisting of 40 items predicted academic achievement about equally well as high school GPA. High-school GPA was a slightly better predictor for third year academic performance, but the curriculum-sampling test was a slightly better predictor for first year progress and retention. It is remarkable that a relatively simple multiple-choice exam performs as well as high school GPA, when we take into account that high school GPA consists of grades obtained on high-school exams across several years combined with grades on national final exams. These findings also support the notion that applying a content-matching approach to predictors and outcomes (e.g., Sackett et al., 2016) is beneficial for predictive validity. The importance of content matching was further supported by the finding that for performance in statistics courses, a math test showed incremental validity over the curriculum-sampling test.

In chapter 4 differential prediction analyses using sample-based assessments were presented. The results showed that curriculum sampling yielded no or little differential prediction by gender, with small effect sizes. In addition, it was shown that increasing the representativeness or comprehensiveness of curriculum samples reduced differential prediction. As discussed in chapter 6, curriculum samples were perceived favorably by applicants for selection and for matching purposes, arguably due to their high-fidelity nature. In addition, as shown in chapters 2 and 3, curriculum-sampling tests scores were also related to enrollment

decisions, and possibly self-selection. Given the main aim of admission procedures in the Netherlands to 'get the right students at the right place', this may be one of the most practically relevant results.

A possible shortcoming of curriculum samples and of simulation-based exercises in general (e.g., Lievens & De Soete, 2012), is that sample-based assessments are black boxes: We do not know what they measure. In chapter 3, we hypothesized that a curriculum sample taps into several cognitive and noncognitive abilities, skills, and traits that are also related to academic performance. The results only partially confirmed these expectations, and some results were contrary to our expectations and difficult to interpret. We did find some noncognitive saturation, which indicated that curriculum sampling may be able to serve as an alternative to self-reports in capturing noncognitive characteristics. On the other hand, results from chapter 5 showed that noncognitive characteristics assessed in a low-stakes condition did add substantial incremental validity over the curriculum-sampling test score. So, there seems to be quite some noncognitive variance that is not captured by the score on the curriculum-sampling test. This is a topic that deserves more attention in future research.

10.1.4 Cognitive Abilities and Skills

As discussed in chapter 1, the strongly cognitively loaded achievement tests like the SAT and the ACT in the USA, are almost never used in European admission procedures. Therefore, they were not used in the empirical studies in this thesis. However, results from chapter 3 based on, admittedly, a relatively small sample showed a non-significant relationship between scores on a cognitive ability test and performance in the first year. Compared to results from the U. S. (Kuncel & Hezlett, 2010; Sackett et al., 2009; Shen et al., 2012), substantially lower correlations between scores on strongly cognitively-loaded tests and academic performance in higher education are commonly found in European studies (Lyren, 2008; Wolming, 1999, Busato et al., 2000; Kappe & van der Flier, 2012). This is probably due to the early selection and stratification of the European education system (Crombag et al., 1975; Resing & Drenth, 2007). Therefore, the applicant population for European higher education may be more homogeneous with respect to cognitive abilities than the applicant population to higher education in the USA. Differential prediction of cognitive ability tests was not studied in this thesis, but results from previous studies showed that female academic performance is typically slightly underpredicted and ethnic minority performance is usually overpredicted by these tests (Fischer et al., 2013; Keiser et al., 2016; Sackett et al., 2008; Wolming, 1999). The findings described in chapter 6 showed

that cognitive ability tests were perceived as moderately favorable for selection and matching procedures.

So, the findings in chapter 3, although based on a relatively small sample, are in line with earlier findings that tests of general cognitive abilities may not be very suitable to differentiate successful from unsuccessful applicants in the highly restricted population of Dutch college applicants (Crombag et al., 1975; Resing & Drenth, 2007). This conclusion may hold to a lesser extent for universities of applied sciences, with their more diverse applicant pool in terms of educational background. Combined with the findings on stakeholder favorability and previous findings about differential prediction, these results do not encourage the use of these measures in admission to European higher education.

10.2 Selective Admission in the Netherlands: Is it worth the trouble?

As discussed in chapter 1, in the Dutch society there is still a debate about admission through assessment versus admission through a (weighted) lottery. According to Stone (2008a, 2008b), the choice between the two systems in terms of fairness depends on whether we have arguments to grant admission to some students over others; Lotteries are fair when students have equally good 'claims' to be admitted. When we define the validity of a student's claim to admission in terms of their future academic achievement, some authors (van der Maas & Visser, 2017; Visser, 2017) have argued that we do not have sufficient arguments to favor admission trough assessment over lottery admission because we are not able to differentiate between applicants in terms of their suitability to study in certain programs. Admission decisions are made based on a rank ordering of applicants through their admission test scores. According to these critics, this rank ordering is invalid, because we cannot claim that, say, the applicant with rank 301 is less suitable than the applicant with rank 300, even when highly reliable and valid admission tests were used. That is correct. Moreover, the same argument can be used in almost all situations in which psychological and educational tests are used to select or assess people. Small differences in test performance between individuals cannot be interpreted as true differences in skills, traits, and abilities, because we cannot measure them *that* precisely. But the conclusion that we do not have arguments to select people does not follow from this reasoning. If we select a set of applicants with the highest ranks based on a reliable and valid procedure, instead of drawing a random sample from the applicant pool, the academic achievement of the selected group will be higher (Dawes, 1979; Taylor & Russel, 1939; Naylor & Shine, 1965). The results in chapters 2 and 3 and in many other studies have shown that there are sufficiently valid measures to predict later academic achievement. So, if we want to select students with the best academic

potential, I conclude that we do have arguments to differentiate between students' 'claims' to admission.

10.2.1 Effects of Selection at the Program Level

However, the effects of admission procedures also depend on the base rate and the selection ratio. When we take these factors into account, the effects of selection by assessment are small or nonexistent in most 'selective' programs in the Netherlands, as discussed in chapter 7. The main limitation of this conclusion is that we assume that the quality of the applicant pool is the same under admission by lottery and admission through assessment. That may not be the case. Implementing an admission procedure and the specific content of that admission procedure may result in changes in the applicant pool, discouraging some to apply at all, and encouraging others. This topic received little attention, but some studies do indicate that such effects exist, and that at least some students base their decision to apply for a study program on the type of admission procedure (see chapter 6; Wouters et al., 2017). Thus, implementing assessments may cause changes in academic achievement of the selected group, even when the selection ratio equals one.

10.2.2. Effects of Selection at a National Level

What is the effect of selective admission procedures on academic achievement in Dutch higher education? First, we should keep in mind that the general aim was to 'get the right student at the right place', while ensuring accessibility to higher education. All students who meet the minimum admission requirements should be able to follow a study program (Wet Kwaliteit in Verscheidenheid, 2013). That means that the selection ratio of applicants to higher education approaches or equals one; College admission in the Netherlands is mostly about allocation, rather than selection. Applicants are not selected for or matched to the level of education, but to a specific program (e.g., psychology, law). That also means that most admission procedures should not be aimed at predicting academic achievement in college, but at predicting academic achievement within a specific college program. Consequently, it does not make sense to include general predictors of academic achievement that are unrelated to the discipline of interest, such as cognitive ability or conscientiousness. A low conscientiousness score would probably lead to rejection or a negative enrollment advice at any program, and is thus not in line with the aims of admission procedures: assessing student-program fit. Also, the overall effects of admission through assessment on dropout rates, time to completion, academic performance and thus costs and resources, are probably strongly overstated. This is nicely illustrated by Cronbach's (1984) response to a

claim that implementing a better personnel selection procedure for programmers would save the American economy a staggering amount of money.

> This projection is a fairytale. The economy utilizes most of the persons who are trained as programmers, and only the most prestigious firms can reject [a substantial percentage] of those who apply. If 90 percent of all programmers are hired somewhere, the tests merely gives a competitive advantage to those firms that test (when some others do not test). Essentially, the benefit would come from routing each person in the labor market into the career where he or she can make the greatest contribution (p. 383).

The same rationale applies to admission to higher education. Admission testing mostly benefits the few programs that can afford to reject a substantial proportion of their applicants, but it will probably hardly affect the quality of the student population in higher education, unless we reject a substantial proportion of the applicants who meet the minimum admission requirements. Perhaps there may be a positive effect of using content-matched admission procedures that promote self-selection and potentially reduce dropout and switching between study programs. Another argument against switching back to lottery admission is that, according to the results in chapter 6, applicants perceived lottery as the least favorable admission method.

Finally, I should note that these analyses entirely depend on the definition of the aim of admission. I assumed that the aim is predicting academic achievement, or getting the right students at the right place. However, when the aim is, for example, to admit a diverse class of students (e.g., Stemler, 2017; Zwick, 2017, pp. 173-183), or a student body that is maximally representative of society (Stegers-Jager et al., 2015), lottery is clearly the most efficient and effective system to meet that aim in the Dutch context.

10.3 Limitations and Future Research

There are limitations to the research presented in this thesis. First, all empirical studies were conducted using samples of applicants to a psychology program. Therefore, the results do not necessarily generalize to applicants to other programs. With respect to the predictive validity of curriculum-sampling tests we expect to find similar results in other predominantly theory-oriented programs. However, it would be more challenging to design curriculum samples for more practically oriented- or vocational programs. The multiple mini-interview approach applied in medicine (Pau et al., 2013; Reiter et al., 2007) or practical skill

simulations (Valli & Johnson, 2007; Vihavainen et al., 2013) may provide good alternatives to exam-based curriculum samples. However, they are less efficient than simple exams because they are more complex to develop and to administer. Predicting academic achievement in admissions to practically oriented programs and vocational education deserves attention in future research. Similarly, it would be valuable to replicate the studies on self-presentation and applicant perceptions derived in chapters 5 and 6 in more heterogeneous applicant samples.

Second, the main focus of this thesis was on admission to selective undergraduate programs. However, the majority of Dutch higher education programs have open admissions with a matching procedure that results in a non-binding enrollment advice. Chapter 6 showed that there were only small differences between applicant perceptions of admission methods in a selection or a matching context. An interesting topic for future research is whether the results on the predictive validity of curriculum samples and content-matched skills tests generalize to low-stakes matching procedures. Furthermore, enrollment decisions and self-selection deserve special attention in this context. Given the widespread use of motivation and personality questionnaires in matching procedures, it would also be interesting to investigate the presence and effects of impression management in a matching context.

Third, differential prediction by ethnicity or SES was not studied in this thesis, but is an important topic for future research. Incorporating ethnic background in research with Dutch samples is not straightforward because the label "ethnic minorities" covers a very heterogeneous group of persons consisting of several relatively small subgroups with different characteristics. It is difficult to obtain samples of sufficient size from each subgroup to do differential prediction analyses. Nevertheless, there is a great need for this type of research, since there are differences for access to (e.g., Stegers-Jager, et al., 2015; van den Broek et al., 2017) and performance (e.g., Meeuwise, Born, & Severiens, 2013, 2014; Ooijevaar, 2010) in higher education between ethnic-minority and ethnic-majority students.

Finally, the outcome measures used to assess academic achievement had some drawbacks. Retention was defined as dropping out of the program, but no distinction was made between switching to another program and dropping out of higher education. The variables GPA and obtained credits have some flaws as well. First, the number of obtained credits was skewed to the left. Second, in the first- and second year, most of the curriculum was fixed, but most courses in the third year were elective. The difficulty and type of courses chosen was not taken into account in this thesis. Also, even in the first year, not all students participated in

every exam due to dropout and study delays. Nevertheless, we computed GPA as the mean grade obtained by each student, which implicitly means that we used the available grades to replace missing data due to courses and exams that were not taken. This probably leads to an underestimation of variance in the outcomes measures, resulting in an underestimation of the strength of the relationships with predictor variables (Smits, 2003; Vickers, 2000). However, despite these shortcomings, GPA tends to be a highly reliable measure (Bacon & Bean, 2006; Beatty et al., 2015).

10.4 Scientific Contributions

An important contribution of the research presented in this thesis is the promising results obtained using curriculum-sampling tests. These results showed that a samples approach can successfully be applied to predict future academic achievement and thus to select applicants in college admissions. Few studies have been conducted on this topic, and even fewer have explicitly made the theoretical link to the samples-approach and the theory of behavioral consistency (exceptions are Lievens & Coetsier, 2002 and Patterson et al., 2012). It is sometimes argued that sample-based assessments and predictors do not contribute to our understanding of performance, that they are atheoretical, and that a correlation between a sample-based assessment score and future performance is not a validity coefficient, but an expression of reliability (e.g., Wernimont & Campbell, 1968). The lack of a foundation in psychological constructs seems to be the main objection. I disagree with this criticism. The theory of behavioral consistency underlying this approach is a theory in itself, be it a simple one that does not require complex theoretical frameworks that specify relations between several unidimensional constructs. It is certainly true that defining and measuring distinct psychological constructs can be a very useful approach to understand, measure, and predict behavior and performance. But when it comes to prediction, defining constructs is a tool, it is not the ultimate goal (e.g., van der Flier 1992). Sijtsma and Evers (2011) stated that this practical focus on *if* something works rather than *why* something works shows a lack of curiosity. Indeed, the choice to study a samples approach was driven by a practical need for a method that predicts well and that taps into a mixture of different psychological constructs. As shown in chapter 3, I was curious about what those constructs might be. This question is, above all, interesting, but may also provide insights to improve the development of sample-based assessments. However, I do not think that answering this question is absolutely necessary to legitimize the use of sample-based assessments. As Baumeister, Vohs, and Funder (2007) discussed, the exclusive focus on constructs can even distract us from the goal of studying *actual* behavior and performance.

In the studies presented in this thesis the aim was to predict future academic achievement. The results showed that this was possible without relying on psychological constructs. It is interesting to speculate about what we would have found if we had adopted sign-based theories and assessments. The results in Chapters 3 and 5 indicated that a combination of unidimensional test scores of, for example cognitive abilities (Kuncel & Hezlett, 2010), personality and motivation questionnaires (Richardson et al., 2012), and study skills and study habits (Credé & Kuncel, 2008), would have yielded lower predictive validity in the present setting than the "atheoretical" samples approach. In some contexts, samples can outperform sign-based assessments in predictive validity, differential prediction, and face validity (e.g., Schmitt in Morgeson et al., 2017b, p. 715). Finally, as noted by van der Flier (1992), the distinction between signs and samples is a relative one, ranging from 'pure' measures of intelligence and personality, measures of 'college readiness' and contextualized behavioral scales, to representative job simulations. Each has its benefits, depending on the context and the aim.

Another important finding in this thesis was that administering self-report questionnaires in a high-stakes context attenuated predictive validity as compared to administering these questionnaires in a low-stakes context. This topic was not studied before in an educational context. The word "context" is important here, because the participants in this study were explicitly informed that in the high-stakes context (i.e., in the admission procedure), their scores would *not* affect admission decisions, and that completing the questionnaire was voluntary and for research purposes only. Nevertheless, a substantial proportion of the respondents were flagged for inflated scores compared to their scores obtained after admission, and the validities of these scores were, in general, lower in the high-stakes condition. This shows that the terms low- and high-stakes should be interpreted with care and that *perceived* stakes and actual stakes may not always coincide.

Finally, little was known about applicant perceptions of methods used for admission to higher education. The study described in chapter 6 showed that, similar to findings obtained in personnel selection procedures, high-fidelity methods and interviews were generally perceived most favorably. However, there were some notable differences. Anderson et al. (2010) found that in personnel selection, methods with the highest predictive validities were also perceived most favorably. The high favorability of the hardly predictive admission interview (Dana et al., 2013; Goho & Blackman, 2006) and the low favorability of strongly predictive high school grades deviate from that finding. In addition, contrary to expectations, we found few differences in favorability depending on the aim (matching or selection) and gender.

10.5 Practical Contributions

The results presented in this thesis contributed to our knowledge of what works in predicting academic achievement in admission procedures and they have implications for admission procedures in practice. First, I advise not to use self-report instruments for selective admission purposes given the findings presented in chapter 5. Second, I advise not to use tests of general cognitive abilities or scholastic achievement. Given the lack of stratification and the lack of national final exams in secondary education it makes sense to use such tests like the SAT and the ACT in the USA, but not in Dutch college admissions, especially for research universities. High school grades are good predictors of academic achievement and are efficient to use. However, considering their practical shortcomings (e.g., not standardized across schools and countries) and the negative applicant perceptions, an admission test in the form of a curriculum sample is a good alternative. With curriculum samples, all applicants are assessed on the same criteria. In addition, the predictive validity of the curriculum-sampling test was high, differential prediction was small, and the applicant perceptions were positive. This approach also showed benefits over high school grades in predicting important outcomes like first year retention and progress.

However, curriculum samples do have some practical restrictions. First, more comprehensive curriculum-samples were better predictors and showed less differential prediction, but would reduce the efficiency of the procedure. Second, as discussed in the personnel selection literature on work samples, samples measure what persons can do, not what they will be able to do after some training or experience. This criticism may be dealt with through allowing or even requiring applicants to prepare or study for curriculum-sampling tests. However, the ability to prepare is also a potential threat to fairness, since some applicants, mostly from wealthier backgrounds, hire tutors and take admission test training. Some admission tests (e.g., the BMAT; Cambridge Assessment, 2017) are even explicitly designed not to require specific preparation for this reason. Indeed, the possibility of seeking paid help in admission test preparation is something to take into account. However, preparation is an intrinsic part of studying, and for that reason, I think preparation should be encouraged in admission testing. Also, it is the question if such test preparation courses actually yield substantial benefits in terms of higher test scores (e.g., Kuncel & Hezlett, 2007). In order to reduce unwanted effects and inequalities in access to test preparation resources, colleges could provide such resources to all potential applicants for free (Stemig et al., 2015).

Since curriculum samples consist of tasks that simulate the future study program, they are most suitable for discipline-specific admission procedures. In addition, each college program would need to develop their own curriculum sample(s), which is less efficient than using admission instruments that can be applied more broadly, like general achievement tests. This need for local development also implies that all programs need to have some expertise of test construction to ensure sufficient (psychometric) quality of the curriculum samples. One possibility to overcome this drawback is that multiple programs in specific disciplines, such as psychology or medicine, collaborate to develop curriculum samples that can be used in admission procedures at different universities. In addition, efficiency aside, the need for discipline specific tests is an advantage: It meets the aim of selecting and matching for an educational program, rather than for an educational level. In short, I think that most of these drawbacks can be addressed. European higher education may even be a particularly suitable context for sample-based assessments. I hope that these findings will help admission officers in their decisions on how to design their admission procedures.

10.6 Concluding Remarks

Currently, the biggest challenge in high-stakes testing in education and in other contexts is the valid measurement of noncognitive traits and skills. The majority of the research efforts on this topic aim to improve existing approaches to measure these skills. Examples are the use of the forced-choice format, conditional reasoning tests, detection warnings, and bogus items. As Schmitt noted in Morgeson et al. (2007b, p. 715) we are essentially trying to find ways to fool our respondents, and to conceal what we measure or what a desired response is. College admissions procedures and criteria should be transparent and explicit (Zwick, 2017), and these approaches reduce transparency. For that same reason, the increasingly popular 'holistic evaluation' aimed at taking the whole person into account in its unique combination of traits and skills based on 'expert judgment' (Highhouse & Kostek, 2013; Warps et al., 2017) does not meet those standards and is therefore not a solution. Although many find this idea sympathetic, it offers opportunities for all sorts of biases (Dawes, 1979). As stated by Steven Pinker (as cited in Zwick, 2017, p. 45), 'anything can be hidden behind the holistic fig leaf". Assessments based on a samples approach may be a viable alternative and deserve a more prominent place within this line of research.

References

Abrahams, N. M., Alf, E. F., & Wolfe, J. J. (1971). Taylor-Russell tables for dichotomous criterion variables. *Journal of Applied Psychology*, *55*, 449-457. doi:10.1037/h0031761

ACT (2014). *National collegiate retention and persistence to degree rates.* Retrieved from:https://www.ruffalonl.com/documents/shared/Papers_and_Research/ACT_Data/ACT_persistence_2014.pdf

Adam, J., Bore, M., Childs, R., Dunn, J., McKendree, J., Munro, D., & Powis, D. (2015). Predictors of professional behaviour and academic outcomes in a UK medical school: A longitudinal cohort study. *Medical Teacher, 37,* 868-880. doi:10.3109/0142159X.2015.1009023

Aguinis, H., Culpepper, S. A., & Pierce, C. A. (2010). Revival of test bias research in preemployment testing. *Journal of Applied Psychology*, *95*, 648-680. doi:10.1037/a0018714

American Educational Research Association, American Psychological Association, & the National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Anderson, N., Salgado, J. F., & Hülsheger, U. R. (2010). Applicant reactions in selection: Comprehensive meta-analysis into reaction generalization versus situational specificity. *International Journal of Selection and Assessment, 18*, 291-304. doi:10.1111/j.1468-2389.2010.00512.x

Anderson, N. & Witvliet, C. (2008). Fairness reactions to personnel selection methods: An international comparison between the Netherlands, the United States, France, Spain, Portugal, and Singapore. *International Journal of Selection and Assessment, 16,* 1-13. doi:10.1111/j.1468-2389.2008.00404.x

Aramburu-Zabala Higuera, L. (2001). Adverse impact in personnel selection: The legal framework and test bias. *European Psychologist*, *6*, 103-111. doi:10.1027//1016-9040.6.2.103

Arneson, J. J., Sackett, P. R., & Beatty, A. S. (2011). Ability-performance relationships in education and employment settings. *Psychological Science, 22,* 1336-1342. doi:10.1177/0956797611417004

Asher, J. J. & Sciarrino, J. A. (1974). Realistic work sample tests: A review. *Personnel Psychology, 27*, 519-533. doi:10.1111/j.1744-6570.1974.tb01173.x

Atkinson, R. C. & Geiser, S. (2009). Reflections on a century of college admissions tests. *Educational Researcher, 38*, 665-676. doi:10.3102/0013189X09351981

Bacon, D. R. & Bean, B. (2006). GPA in research studies: An invaluable but neglected opportunity. *Journal of Marketing* Education, *28,* 35–42. doi:10.1177/0273475305284638

Balf, T. (2014, March 6). The story behind the SAT overhaul. *The New York Times.* Retrieved from: http://nyti.ms/1cCH2Dz

Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology*, *81*, 261–272. doi:10.1037/0021-9010.81.3.261

Barry, C. L., & Finney, S. J. (2016). Modeling change in effort across a low-stakes testing session: A latent growth curve modeling approach. *Applied Measurement in Education*, *29*, 46-64. doi:10.1080/08957347.2015.1102914

Bartels, M., Rietveld, M. H., Van Baal, G. M., & Boomsma, D. I. (2002). Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. *Twin Research*, *5*, 544-553. doi:10.1375/136905202762342017

Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P., & Campion, M. A. (2001). Applicant reactions to selection: Development of the selection procedural justice scale (SPJS). *Personnel Psychology*, *54*, 388-420. doi:10.1111/j.1744-6570.2001.tb00097.x

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*, 396–403. doi:10.1111/j.1745-6916.2007.00051.x

Beatty, A. S., Barratt, C. L., Berry, C. M., & Sackett, P. R. (2014). Testing the generalizability of indirect range restriction corrections. *Journal of Applied Psychology*, *99*, 587-598. doi:10.1037/a0036361

Beatty, A. S., Walmsley, P. T., Sackett, P. R., & Kuncel, N. R. (2015). The reliability of college grades. *Educational Measurement: Issues and Practice, 34*, 31–40. doi:10.1111/emip.12096

Becker, R., & Kolster, R. (2012). *International student recruitment: Policies and developments in selected countries*. The Hague, The Netherlands: Netherlands Organisation for International Cooperation in Higher Education. Retrieved from https://www.epnuffic.nl/en/publications/find-a-publication/international-student-recruitment.pdf

Berry, C. M. (2015). Differential validity and differential prediction of cognitive ability tests: Understanding test bias in the employment context. *Annual Review of Organizational Psychology and Organizational Behavior, 2,* 435–463. doi:10.1146/annurev-orgpsych-032414-111256

Bertolino, M., & Steiner, D. D. (2007). Fairness reactions to selection methods: An Italian study. *International Journal of Selection and Assessment*, *15*, 197-205. doi:10.1111/j.1468-2389.2007.00381.x

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment, 14,* 317–335. doi:10.1111/j.1468–2389.2006.00354.x.

Booij, A. S. & van Klaveren, C. (2017, June). *Trial lectures or admission talks? How to improve students' choice of major.* Paper presented at the Onderwijs Research Dagen [Education Research Days], Antwerp, Belgium.

Borghans, L., Golsteyn, B. H., Heckman, J., & Humphries, J. E. (2011). Identification problems in personality psychology. *Personality & Individual Differences*, *51*, 315-320. doi:10.1016/j.paid.2011.03.029

Borghans, L., Golsteyn, B. H., Heckman, J. J., & Humphries, J. E. (2016). What grades and achievement tests measure. *PNAS Proceedings of the National Academy of Sciences of The United States of America*, *113*, 13354-13359. doi:10.1073/pnas.1601135113

Bouma, K. (2017). Studies met numerus fixus zeer populair: twee keer zo veel aanmeldingen als plekken [Selective programs very popular: Two applicants for each slot]. *De Volkskrant*. Retrieved from: https://www.volkskrant.nl/binnenland/studies-met-numerus-fixus-zeer-populair-twee-keer-zo-veel-aanmeldingen-als-plekken

Bowen, W. G., Chingos, M. M., & McPherson, M. S. (2011). *Crossing the finish line: Completing college at America's public universities*. Princeton, NJ: Princeton University Press.

Bowler, J. L., & Bowler, M. C. (2014). Evaluating the fakability of a conditional reasoning test of addiction proneness. *International Journal of Psychology, 49,* 415–419. doi:10.1002/ijop.12030.

Breaugh, J. A. (2008). Employee recruitment: Current knowledge and important areas for future research. *Human Resource Management Review, 18*, 103-118. doi:10.1016/j.hrmr.2008.07.003

Brody, N. (2003). Construct validation of the Sternberg Triarchic Abilities Test: Comment and reanalysis. *Intelligence*, *31*, 319–329. doi:10.1016/S0160-2896(01)00087-3

Brogden, H. E. (1949). When testing pays off. *Personnel Psychology, 2*, 171-185. doi:10.1111/j.1744-6570.1949.tb01397.x

Brown, A. (2016, July). *Response distortions in self-reported and other-reported measures: is there light at the end of the tunnel?* Paper presented at the 10th International Test Commission Conference, Vancouver, Canada.

Bruni, F. (2016, March 5). Hidden gold in college applications. *The New York Times*. Retrieved from: https://www.nytimes.com/2016/03/06/opinion/sunday/hidden-gold-in-college-applications.html?

Burns, G. N., Fillipowski, J. N., Morris, M. B., & Shoda, E. A. (2015). Impact of electronic warnings on online personality scores and test-taker reactions in an applicant simulation. *Computers in Human Behavior, 48*, 163-172. doi:10.1016/j.chb. 2015.01.051.

Busato, V. V., Prins, F. J., Elshout, J. J., & Hamaker, C. (2000). Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education. *Personality and Individual Differences, 29,* 1057-1068. doi:10.1016/S0191-8869(99)00253-6

Callinan, M. & Robertson, I. T. (2000). Work sample testing. *International Journal of Selection and Assessment*, *8*, 248-260. doi:10.1111/1468-2389.00154

Camara, W. J. (2009). College admission testing: Myths and realities in an age of admissions hype. In R. P. Phelps (Ed.), *Correcting fallacies about educational and psychological testing* (pp. 147–180). Washington, DC: American Psychological Association. doi:10.1037/11861-004

Camara, W. (2013). Defining and measuring college and career readiness: A validation framework. *Educational Measurement: Issues and Practice, 32*, 16–27. doi:10.1111/emip.12016

Cambridge Assessment (2017). *The BioMedical Admissions Test (BMAT).* Retrieved from: http://www.admissionstestingservice.org/for-test-takers/bmat/

Cascio, W. F. (1980). Responding to the demand for accountability: A critical analysis of three utility models. *Organizational Behavior & Human Performance*, *25*, 32-45. doi:10.1016/0030-5073(80)90024-0

Cavagnaro, L. B., & Fasihuddin, H. (2016). A moonshot approach to change in higher education: Creativity, innovation, and the redesign of academia. *Liberal Education, 102*. Retrieved from: https://www.aacu.org/liberaleducation/2016/spring/cavagnaro

Chamorro-Premuzic, T., & Furnham, A. (2003). Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of Research in Personality, 37,* 319–338. doi:10.1016/S0092-6566(02)00578-0

Chamorro-Premuzic, T. & Furnham, A. (2005). *Personality and intellectual competence*. Mahwah, NJ: Lawrence Erlbaum.

Chan, D., Schmitt, N., DeShon, R. P., Clause, C. S., & Delbridge, K. (1997). Reactions to cognitive ability tests: The relationships between race, test performance, face validity perceptions, and test-taking motivation. *Journal of Applied Psychology*, *82*, 300-310. doi:10.1037/0021-9010.82.2.300

Chan, D., Schmitt, N., Sacco, J. M., & DeShon, R. P. (1998). Understanding pretest and posttest reactions to cognitive ability and personality tests. *Journal of Applied Psychology*, *83*, 471-485. doi:10.1037/0021-9010.83.3.471

Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance*, *18*, 267-307. doi:10.1207/s15327043hup1803_4

Cleary, T. A. (1968). Test bias: Prediction of grades of negro and white students in integrated colleges. *Journal of Educational Measurement, 5*, 115–124. doi:10.1111/j.1745-3984.1968.tb00613.x

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates

Credé, M. & Kuncel, N. R. (2008). Study habits, skills, and attitudes: The third pillar supporting collegiate performance. *Perspectives on Psychological Science, 3*, 425-453. doi:10.1111/j.1745-6924.2008.00089.x

Cremonini, L., Leisyte, L., Weyer, E., & Vossensteyn, J. J. (2011). *Selection and matching in higher education: An international comparative study.* Enschede, the Netherlands: Center for Higher Education Policy Studies (CHEPS).

Crombag, H. F., Gaff, J. G., & Chang, T. M. (1975). Study behavior and academic performance. *Tijdschrift voor Onderwijsresearch, 1*, 3-14.

Cronbach, L. J. (1984). *Essentials of psychological testing*. New York, NY: Harper & Row Publishers Inc.

Cronbach, L. J., & Gleser, G. C. (1965). *Psychological tests and personnel decisions.* Urbana, IL: University of Illinois Press.

Dahlke, J. A., & Sackett, P. R. (2017). The relationship between cognitive-ability saturation and subgroup mean differences across predictors of job performance. *Journal of Applied Psychology*. Advance online publication. doi:10.1037/apl0000235

Dana, J., Dawes, R., & Peterson, N. (2013). Belief in the unstructured interview: The persistence of an illusion. *Judgment & Decision Making*, *8*, 512-520.

Dawes, R.M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34*, 571-582. doi:10.1037/0003-066X.34.7.571

de Bruyne, H. C. M. & Mellenbergh, G. J. (1973). Selectie of aanpassing: Kritische kanttekeningen bij het rapport "Selectie voor en in het hoger onderwijs, een probleemanalyse" van prof. dr. A. D. de Groot [Selection or adaptation: Critical notes about the report "Selection for and in higher education, a problem analysis" by prof. dr. A. D. de Groot]. *Universiteit en Hogeschool, 20,* 111- 132.

de Groot, A. D. (1972). *Selectie voor en in het hoger onderwijs: een probleemanalyse* [Selection for and in higher education: A problem analysis]. The Hague, the Netherlands: Staatsuitgeverij.

de Leng, W. E., Stegers-Jager, K. M., Husbands, A., Dowell, J. S., Born, M. P., & Themmen, A. N. (2016). Scoring method of a situational judgment test: Influence on internal consistency reliability, adverse impact and correlation with personality? *Advances in Health Sciences Education, 22*, 243–265. doi:10.1007/s10459-016-9720-7

de Raad, B. & Schouwenburg, H. C. (1996). Personality in learning and education: A review. *European Journal of Personality, 10*, 303-336. doi:10.1002/(SICI)1099-0984

de Ridder, D. (2017, January 24). Je bent jong, lakt je nagels en studeert wat [You are young, you paint your nails, and you study a little]. *NRC Handelsblad.* Retrieved from: https://www.nrc.nl/nieuws/2017/01/24/je-bent-jong-lakt-je-nagels-en-studeert-wat-a1542670

de Visser, M., Fluit, C., Fransen, J., Latijnhouwers, M., Cohen-Schotanus, J., & Laan, R. (2017). The effect of curriculum sample selection for medical school. *Advances in Health Science Education, 22*, 43–56. doi:10.1007/s10459-016-9681-x

Dean, M. A., Roth, P. L., & Bobko, P. (2008). Ethnic and gender subgroup differences in assessment center ratings: A meta-analysis. *Journal of Applied Psychology*, *93*, 685-691. doi:10.1037/0021-9010.93.3.685

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*, 13-21. doi:10.1016/j.intell.2006.02.001

Denissen, J. A., Geenen, R., van Aken, M. G., Gosling, S. D., & Potter, J. (2008). Development and validation of a Dutch translation of the Big Five Inventory (BFI). *Journal of Personality Assessment, 90*, 152-157. doi:10.1080/00223890701845229

Dienst Uitvoering Onderwijs (2014). *Jaarverslag numerus fixus-opleidingen.* [Annual report numerus fixus programs] Retrieved from: https://duo.nl/Images/Jaarverslag-numerus-fixi-studiejaar-2014-2015_tcm7-50219.pdf

Dollinger, S. J. (2011). Standardized minds or individuality? Admissions tests and creativity revisited. *Psychology of Aesthetics, Creativity, and the Arts*, *5*, 329–341. doi:10.1037/a0023659

Downs, S., Farr, R. M., & Colbeck, L. (1978). Self-appraisal: A convergence of selection and guidance. *Journal of Occupational Psychology, 51*, 271-278.

doi:10.1111/j.2044-8325.1978.tb00423.x
Drenth, P. J. D. (1995, March 30). Duijkerlezing; in Nederland is selectie onmogelijk [Duijker lecture: Selective admission is impossible in the Netherlands]. *NRC Handelsblad*. Retrieved from: http://www.nrc.nl/handelsblad/van/1995/maart/30/duijkerlezing-in-nederland-is-selectie-onmogelijk-7262098
Drenth, P. J. D. (1998). Selectie voor en in de studie geneeskunde [Selection for and in medical education]. *Bulletin Medisch Onderwijs, 17*, 97-107.
Dumfart, B., & Neubauer, A. C. (2016). Conscientiousness is the most powerful noncognitive predictor of school achievement in adolescents. *Journal of Individual Differences*, *37*, 8-15. doi:10.1027/1614-0001/a000182
Dwight, A. A., & Donovan, J. J. (2003). Do warnings not to fake reduce faking? *Human Performance*, *16*, 1–23. doi:10.1207/S15327043HUP1601_1
Dynarski, S. (2017, July 14). Simple way to help low-income students: make everyone take SAT or ACT. *The New York Times.* Retrieved from: https://www.nytimes.com/2017/07/14/upshot/how-universal-college-admission-tests-help-low-income-students.html
Education Council. (2014). *Overgangen in het onderwijs* [Transitions in education]. Retrieved from: http://www.onderwijsraad.nl/publicaties/2014/overgangen-in-het-onderwijs/volledig/item7086
Egberink, I. J. L., Holly-Middelkamp, F. R., & Vermeulen, C .S. M. (2010). *Q1000 Capaciteiten Hoog* [COTAN review 2010, Q1000 High Capacities]. Retrieved from: www.cotandocumentatie.nl
Eva, K. W. (2003). On the generality of specificity. *Medical Education, 37*, 587-588. doi:10.1046/j.1365-2923.2003.01563.x
Eva, K. W., Reiter, H. I., Rosenfeld, J., & Norman, G. R. (2004). The ability of the multiple mini-interview to predict pre-clerkship performance in medical school. *Academic Medicine*, *79*, 40-42. doi:10.1097/00001888-200410001-00012
Eva, K. W., Reiter, H. I., Trinh, K., Wasi, P., Rosenfeld, J., & Norman, G. R. (2009). Predictive validity of the multiple mini-interview for selecting medical trainees. *Medical Education, 43,* 767-775. doi:10.1111/j.1365-2923.2009.03407.x
Eva, K. W., Rosenfeld, J., Reiter, H.I., & Norman, G. R. (2004). An admissions OSCE: the multiple mini-interview. *Medical Education, 38,* 314-326. doi:10.1046/j.1365-2923.2004.01776.x
Field, A. P. (2013). *Discovering Statistics Using SPSS*. London, UK: Sage Publications
Fife, D. (2016). *Selection: Correcting biased estimates under selection*. R package version 1.0. Retrieved from: http://CRAN.R-project.org/package=selection
Fischer, F. T., Schult, J., & Hell, B. (2013). Sex-specific differential prediction of college admission tests: A meta-analysis. *Journal of Educational Psychology*, *105*, 478-488. doi:10.1037/a0031956
Furnham, A., Chamorro-Premuzic, T., & McDougall, F. (2003). Personality, cognitive ability, and beliefs about intelligence as predictors of academic performance. *Learning and Individual Differences*, *14*, 47-64. doi:10.1016/j.lindif.2003.08.002

Gardner, D. G., Cummings, L. L., Dunham, R. B., & Pierce, J. L. (1998). Single-item versus multiple-item measurement scales: an empirical comparison. *Educational and Psychological Measurement, 58*, 898-915. doi:10.1177/0013164498058006003

Geiser, S., & Studley, R. (2002). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment*, *8*, 1-26. doi:10.1207/S15326977EA0801_01

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian Data Analysis* (3rd ed.). Boca Raton, FL: CRC Press.

Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review, 18,* 694-734. doi:10.5465/AMR.1993.9402210155

Gilliland, S. W. (1994). Effects of procedural and distributive justice on reactions to a selection system. *Journal of Applied Psychology*, *79*, 691-701. doi:10.1037/0021-9010.79.5.691

Gilliland, S. W. (1995). Fairness from the applicant's perspective: Reactions to employee selection procedures. *International Journal of Selection and Assessment*, *3*, 11-19. doi:10.1111/j.1468-2389.1995.tb00002.x

Goho, J., & Blackman, A. (2006). The effectiveness of academic admission interviews: An exploratory meta-analysis. *Medical Teacher*, *28*, 335-340. doi:10.1080/01421590600603418

Goldstein, H. W., Zedeck, S., & Goldstein, I. L. (2002). g: Is this your final answer? *Human Performance, 15,* 123-142, doi:10.1080/08959285.2002.9668087

Gottfredson, L. S. (2003a). Discussion: On Sternberg's "reply to Gottfredson". *Intelligence, 31*, 415–424. doi:10.1016/S0160-2896(03)00024-2

Gottfredson, L. S. (2003b). Dissecting practical intelligence theory: Its claims and evidence. *Intelligence, 31*, 343–397. doi:10.1016/S0160-2896(02)00085-5

Greiff, S., Martin, R., & Spinath, B. (2014). Introduction to the special section on computer-based assessment of cross-curricular skills and processes. *Journal of Educational Psychology, 106*, 605-607. doi:10.1037/a0035607

Griffin, P., & Care, E. (Eds.). (2015). *Assessment and teaching of 21st century skills: Methods and approach*. Heidelberg, Germany: Springer. doi:10.1007/978-94-017-9395-7

Griffin, B., & Wilson, I. G. (2012). Faking good: Self-enhancement in medical school applicants. *Medical Education, 46,* 485–490. doi:10.1111/j.1365-2923.2011.04208.x.

Griffith, R. L., & Peterson, M. H. (2008). The failure of social desirability measures to capture applicant faking behavior. *Industrial and organizational psychology: Perspectives on Science and Practice, 1*, 308–311. doi:10.1111/j.1754-9434.2008.00053.x.

Grofman, B., & Merrill, S. (2004). Anticipating likely consequences of lottery-based affirmative action. *Social Science Quarterly*, *85*, 1447-1468. doi:10.1111/j.0038-4941.2004.00285.x

Guion, R. M. (1998*). Assessment, measurement, and prediction for personnel decisions.* Mahwah, NJ: Erlbaum

Gulliksen, H. (1950). *Theory of mental tests*. New York, NY: John Wiley & Sons.

Haggerty, M. E. (1918). Tests of applicants for admission to University of Minnesota Medical School. *Journal of Educational Psychology*, *9*, 278-286. doi:10.1037/h0071405

Häkkinen, I. (2004). *Essays on school resources, academic achievement, and student employment* (doctoral dissertation). Uppsala, Sweden: Uppsala University. Retrieved from: http://www.diva-portal.org/smash/get/diva2:165057/FULLTEXT01.pdf

Harris, B. H., Walsh, J. L., & Lammy, S. (2015). UK medical selection: Lottery or meritocracy? *Clinical Medicine, 15*, 566. doi:10.7861/clinmedicine.15-6-566

Hartman, N. S., & Grubb,W. I. (2011). Deliberate faking on personality and emotional intelligence measures. *Psychological Reports, 108,* 120–138. doi:10.2466/03.09.28.PR0.108.1.120-138.

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, *57*, 639-683. doi:10.1111/j.1744-6570.2004.00003.x

Hayes, A. F. & Montoya, A. K. (2017). A tutorial on testing, visualizing, and probing an interaction involving a multicategorical variable in linear regression analysis. *Communication Methods and Measures, 11,* 1-30, doi:10.1080/19312458.2016.1271116

Heine, C., Briedis, K., Didi, H.J. Haase, K., & Trost, G. (2006). *Auswahl- und Eignungsfeststellungsverfahren beim Hochschulzugang in Deutschland und ausgewählten Ländern* [Selection and suitability assessment procedures for university entrance in Germany and some other countries]. Hannover, Germany: Hochschul-Informations-System/ITB Consulting. Retrieved from: http://www.dzhw.eu/pdf/pub_kia/kia200603.pdf

Heggestad, E. D., Morrison, M., Reeve, C. L., & McCloy, R. A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology, 91*, 9–24. doi:10.1037/0021-9010.91.1.9

Hermelin, E., Lievens, F., & Robertson, I. T. (2007). The validity of assessment centers for the prediction of supervisory performance ratings: A meta-analysis. *International Journal of Selection and Assessment, 15*, 405–411. doi:10.1111/j.1468-2389.2007.00399.x

Highhouse, S., & Kostek, J. A. (2013). Holistic assessment for selection and placement. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J. C. Hansen, N. R. Kuncel, S. P. Reise, ... M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology* (pp. 565-577). Washington, DC, US: American Psychological Association. doi:10.1037/14047-031

Hirsh, J. B. & Peterson, J. B. (2008). Predicting creativity and academic success with a "fake-proof" measure of the big five. *Journal of Research in Personality, 42*, 1323–1333. doi:10.1016/j.jrp.2008.04.006.

Hofstee, W. K. B. (1972). De Groot komt met beperkte maatregelen, niet met een creatief beleid [De Groot proposes limited measures instead of creative policies]. *De Groene*, 2-10.

Holling, H. (1998). Utility analysis of personnel selection: An overview and empirical study based on objective performance measures. *Methods of Psychological Research, 3*, 5-24.

Hook, K. M., & Pfeiffer, C. A. (2007). Impact of a new curriculum on medical students' interpersonal and interviewing skills. *Medical Education, 41*, 154-159. doi:10.1111/j.1365-2929.2006.02680.x

Hoover, E. (2013). Noncognitive measures: The next frontier in college admissions. *Chronicle of Higher Education*. Retrieved from: https://www.collegesuccessfoundation.org/document.doc?id=851

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment*, *9*, 152-194. doi:10.1111/1468-2389.00171

Hox, J. J., Moerbeek, M., & van der Schoot, R. (2010). *Multilevel analysis: Techniques and applications* (2nd ed). New York, NY: Routledge

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage Publications Inc.

Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology*, *91*, 594-612. doi:10.1037/0021-9010.91.3.594

Huws, N., Reddy, P. A., & Talcott, J. B. (2009). The effects of faking on non-cognitive predictors of academic performance in university students. *Learning and Individual Differences, 19*, 476–480. doi:10.1016/j.lindif.2009.04.003

Inspectorate of Education (2017). *Selectie: meer dan cijfers alleen* [Selection: more than just numbers]. The Hague, the Netherlands: Ministry of Education, Culture, and Science.

ISO (2014). *Meer transparantie bij decentrale selectie: Het belang en een framework* [More transparancy in admissions: The importance of a framework]. Retrieved from: http://www.iso.nl/website/wp-content/uploads/2014/12/1415-meer-transparantie-bij-decentrale-selectie3.pdf

Ispas, D., Ilie, A., Iliescu, D., Johnson, R. E., & Harris, M. M. (2010). Fairness reactions to selection methods: A Romanian study. *International Journal of Selection and Assessment*, *18*, 102-110. doi:10.1111/j.1468-2389.2010.00492.x

James, L. R. (1998). Measurement of personality via conditional reasoning. *Organizational Research Methods, 1,* 131–163. doi:10.1177/109442819812001.

Jencks, C. (1998). Racial bias in testing. In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap* (pp. 55–85). Washington, DC: Brooking Institute Press.

John, O. P., Donahue, E.M., & Kentle, R. L. (1991). *The big five inventory–versions 4a and 54*. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

Johns, G. (2006). The essential impact of context of organization behavior. *Academy of Management Review, 31,* 366-408. doi:10.5465/AMR.2006.20208687

Johnson, P. O., & Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems. *Statistical Research Memoirs, 1,* 57–93.
Jordan, J. S. & Turner, B. A. (2008). The feasibility of single-item measures for organizational justice. *Measurement in Physical Education and Exercise Science, 12*, 237–257. doi:10.1080/10913670802349790
Julian, E. R. (2005). Validity of the medical college admission test for predicting medical school performance. *Academic Medicine, 80,* 910-917. doi:10.1097/00001888-200510000-00010
Kappe, R. & van der Flier, H. (2012). Predicting academic success in higher education: What's more important than being smart? *European Journal of Psychology of Education, 27*, 605-619. doi; 10.1007/s10212-011-0099-9
Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90* , 773–795. doi:10.1080/01621459.1995.10476572
Kaufman, J. C. (2010). Using creativity to reduce ethnic bias in college admissions. *Review of General Psychology*, *14*, 189–203. doi:10.1037/a0020133
Keiser, H. N., Sackett, P. R., Kuncel, N. R., & Brothen, T. (2016). Why women perform better in college than admission scores would predict: Exploring the roles of conscientiousness and course-taking patterns. *Journal of Applied Psychology*, *101*, 569-581. doi:10.1037/apl0000069
Kiefer, C., & Benit, N. (2016).What is applicant faking behavior? A review on the current state of theory and modeling techniques. *Journal of European Psychology Students, 7*, 9–19. doi:10.5334/jeps.345.
Klehe, U., Kleinmann, M., Hartstein, T., Melchers, K. G., König, C. J., Heslin, P. A., & Lievens, F. (2012). Responding to personality tests in a selection context: The role of the ability to identify criteria and the ideal-employee factor. *Human Performance, 25,* 273–302. doi:10.1080/08959285.2012.703733.
Kleijn, W. C., van der Ploeg, H. M., & Topman, R. M. (1994). Cognition, study habits, test anxiety, and academic performance. *Psychological Reports*, *75*, 1219-1226. doi:10.2466/pr0.1994.75.3.1219
Kluger, A. N., & Rothstein, H. R. (1993). The influence of selection test type on applicant reactions to employment testing. *Journal of Business and Psychology*, *8*, 3-25. doi:10.1007/BF02230391
Korthals, A.H. (2007). *Commissie 'Ruim baan voor talent'. Eindrapportage* [Committee "Room for talent" final report]. The Hague, the Netherlands: Ministry of Education, Culture, and Science. Retrieved from: https://www.rijksoverheid.nl/actueel/nieuws/2007/12/11/eindrapport-commissie-ruim-baan-voor-talent
Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods, 15,* 722-752. doi:10.1177/1094428112457829.
Kruschke, J. K. & Liddell, T. M. (2017). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review.* Advance online publication. doi:10.3758/s13423-016-1221-4
Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science*, *315*, 1080-1081. doi:10.1126/science.1136618

Kuncel, N. R. & Hezlett, S. A. (2010). Fact and fiction in cognitive ability testing for admissions and hiring decisions. *Current Directions in Psychological Science, 19*, 339-345. doi:10.1177/0963721410389459

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychollogical Buletin, 127,* 162–181. doi:10.1037/0033-2909.127.1.162

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all*? Journal of Personality and Social Psychology, 86,* 148–161. doi:10.1037/0022-3514.86.1.148

Kunina, O., Wilhelm, O., Formazin, M., Jonkmann, K., & Schroeders, U. (2007). Extended criteria and predictors in college admission: Exploring the structure of study success and investigating the validity of domain knowledge. *Psychology Science*, *49*, 88-114.

Kyllonen, P. (2017, July). *Use of forced-choice assessments for higher education admissions.* Paper presented at the 14th European Conference on Psychological Assessment, Lisbon, Portugal.

Kyllonen, P. & Bertling, J. (2017, April). *Interpersonal and intrapersonal skills assessment: Design, development, scoring, and reporting.* Workshop provided at the Annual Meeting of the National Council on Measurement in Education, San Antonio, Texas.

Kyllonen, P. C., Lipnevich, A. A., Burrus, J., & Roberts, R. D. (2014). Personality, motivation, and college readiness: A prospectus for assessment and development. *ETS Research Report Series*, 1-48. doi:10.1002/ets2.12004

Kyllonen, P. C., Walters, A. M., & Kaufman, J. C. (2005). Noncognitive constructs and their assessment in graduate education: A review. *Educational Assessment*, *10*, 153-184. doi:10.1207/s15326977ea1003_2

Lautenschlager, G. J., & Mendoza, J. L. (1986). A step-down hierarchical multiple regression analysis for examining hypotheses about test bias in prediction. *Applied Psychological Measurement, 10,* 133–139. doi:10.1177/014662168601000202

Lawshe, C. H., Bolda, R. A., Brune, R. L., & Auclair, G. (1958). Expectancy charts II: Their theoretical development. *Personnel Psychology*, *11,* 545-559. doi:10.1111/j.1744-6570.1958.tb00040.x

Lay, C. H. (1986). At last, my research article on procrastination. *Journal of Research in Personality*, *20*, 474-495. doi:10.1016/0092-6566(86)90127-3

Lemann, N. (1999). *The big test: The secret history of the American meritocracy*. New York, NY: Farrar, Straus and Giroux.

Lievens, F. (2013). Adjusting medical school admission: Assessing interpersonal skills using situational judgement tests. *Medical Education*, *47*, 182-189. doi:10.1111/medu.12089

Lievens, F. & Coetsier, P. (2002). Situational tests in student selection: An examination of predictive validity, adverse impact, and construct validity. *International Journal of Selection and Assessment,*10, 245–257. doi:10.1111/1468-2389.00215

Lievens, F., & De Soete, B. (2012). Simulations. In N. Schmitt (Ed.) , *The Oxford handbook of personnel assessment and selection* (pp. 383-410). New York, NY: Oxford University Press. doi:10.1093/oxfordhb/9780199732579.013.0017

Lievens, F., & Sackett, P. R. (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology*, *97*, 460-468. doi:10.1037/a0025741

Lievens, F. & Sackett, P. R. (2017). The effects of predictor method factors on selection outcomes: A modular approach to personnel selection procedures*. Journal of Applied Psychology, 102,* 43-66. doi:10.1037/apl0000160

Lievens, F., Sackett, P. R., & Buyse, T. (2009). The effects of response instructions on situational judgment test performance and validity in a high-stakes context. *Journal of Applied Psychology, 94*, 1095–110. doi:10.1037/a0014628

Lyons, B. D., Hoffman, B. J., Michel, J. W., & Williams, K. J. (2011). On the predictive efficiency of past performance and physical ability: The case of the National Football League. *Human Performance*, *24*, 158–172. doi:10.1080/08959285.2011.555218

Lyren, P. (2008). Prediction of academic performance by means of the Swedish scholastic assessment test. *Scandinavian Journal of Educational Research*, *52*, 565-581. doi:10.1080/00313830802497158

Macan, T. H., Avedon, M. J., Paese, M., & Smith, D. E. (1994). The effects of applicants' reactions to cognitive ability tests and an assessment center. *Personnel Psychology*, *47*, 715-738. doi:10.1111/j.1744-6570.1994.tb01573.x

MacKenzie, R .K., Dowell, J., Ayansina, D., & Cleland, J. A. (2017). Do personality traits assessed on medical school admission predict exit performance? A UK-wide longitudinal cohort study. *Advances in Health Science Education, 22*, 365-385. doi:10.1007/s10459-016-9715-4

Mattern, K. D. & Patterson, B. F. (2013). Test of slope and intercept bias in college admissions: A response to Aguinis, Culpepper, and Pierce (2010). *Journal of Applied Psychology*, 98, 134-147. doi:10.1037/a0030610

Mattern, K. D., Patterson, B. F., Shaw, E. J. Kobrin, J. L., & Barbuti, S. M. (2008). *Differential validity and prediction of the SAT* (Research Report No. 2008-4). Retrieved from: https://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2008-4-differential-validity-prediction-sat.pdf

Mattern, K. Sanchez, E., & Ndum, E. (2017). Why do achievement measures underpredict female academic performance? *Educational Measurement: Issues and Practice, 36*, 47-57. doi:10.1111/emip.12138

Maxwell, S. E. & Arvey, R. D. (1993). The search for predictors with high validity and low adverse impact: Compatible or incompatible goals? *Journal of Applied Psychology*, *78*, 433–437. doi:10.1037/0021-9010.78.3.433

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. I. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology, 60*, 63-91. doi:10.1111/j.1744-6570.2007.00065.x

McDaniel, M. A., & Whetzel, D. L. (2005). Situational judgment test research: Informing the debate on practical intelligence theory. *Intelligence*, *33*, 515–525. doi:10.1016/j.intell.2005.02.001

McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology, 85*, 812–821. doi:10.1037//0021-9010.85. 5.812.

McLellan, R. A. (1996). Theoretical expectancy calculator [online utility calculator]. Retrieved from: http://www.hr-software.net/cgi/TheoreticalExpectancy.cgi

Meade, A. M. & Fetzer, M. (2009). Test bias, differential prediction, and a revised approach for determining the suitability of a predictor in a selection context. *Organizational Research Methods, 12,* 738 -761. doi:10.1177/1094428109331487

Meeuwisse, M., Born, M. P., & Severiens, S. E. (2013). Academic performance differences among ethnic groups: Do the daily use and management of time offer explanations? *Social Psychology of Education, 16*, 599-615. doi:10.1007/s11218-013-9231-9

Meeuwisse, M., Born, M. P., & Severiens, S. E. (2014). The family-study interface and academic outcomes: Differences and similarities between ethnic minority and ethnic majority students. *Cultural Diversity and Ethnic Minority Psychology, 20,* 401-412. doi:10.1037/a0036698

Meijer, R. R., & Niessen, A. S. M. (2015). A trial-studying approach to predict college achievement. *Frontiers in Psychology, 6*, 1–3. doi:10.3389/fpsyg.2015. 00887.

Meijler, F. L., & Vreeken, J. (1975, January 17). Lottery admissions system in the Netherlands (Letter to the editor). *Science*, p. 114.

Mellenbergh, G. J. (1993). Beslissen met tests en studietoetsen [Decision-making with tests and educational exams] In: Koele, P. & van der Pligt, J. *Beslissen en beoordelen* [Deciding and assessing] (pp. 96-124). Amsterdam, the Netherlands: Boom

Mellenbergh, G. J. (1995). Selectie aan de poort [Selection at the gate]. *Spiegeloog*, 8-9.

Merckelbach, H. (2015, February 28). Wie bluft mag dokter worden [Those who bluff get to be a doctor]. *NRC Handelsblad.* Retrieved from: https://www.nrc.nl/nieuws/2015/02/28/wie-bluft-mag-dokter-worden-1470326-a341867

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York, NY: Macmillan Publishing Co, Inc.

Miles, J. & Shevlin, M. (2001). *Applying regression & correlation: A guide for students and researchers.* London, U.K.: Sage

Ministry of Education, Culture, and Science (2014, August 29). *Informatie over de afschaffing van loting bij numerusfixusopleidingen* [Information about the abolishment of lottery admission]. [Letter of government]. Retrieved from: http://www.rijksoverheid.nl/documenten-en-

publicaties/kamerstukken/2014/08/30/kamerbrief-met-informatie-over-de-afschaffing-van-loting-bij-numerusfixusopleidingen.html
Moneta-Koehler L., Brown, A. M., Petrie, K. A., Evans, B. J., & Chalkley, R. (2017). The limitations of the GRE in predicting success in biomedical graduate school. *PLoS ONE 12*(1): e0166742. doi:10.1371/journal.pone.0166742
Morey, R. D. & Rouder, J. N. (2015). BayesFactor: Computation of Bayes factors for common designs. R package version 0.9.12-2. Available from: http://CRAN.R-project.org/package=BayesFactor
Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007a). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology, 60*, 1029–1049. doi:10.1111/j.1744-6570.2007.00100.x.
Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007b). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60*, 683–729. doi:10.1111/j.1744-6570.2007.00089.x.
Moscoso, S., & Salgado, J. F. (2004). Fairness reactions to personnel selection techniques in Spain and Portugal. *International Journal of Selection and Assessment*, *12*, 187-196. doi:10.1111/j.0965-075X.2004.00273.x.
Moutafi, J., Furnham, A., & Paltiel, L. (2004). Why is conscientiousness negatively correlated with intelligence? *Personality and Individual Differences*, *37*, 1013–1022. doi:10.1016/j.paid.2003.11.010
Mueller-Hanson, R., Heggestad, E. D., & Thornton, G. I. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology, 88,* 348–355. doi:10.1037/0021-9010.88.2.348.
Murphy, S. C., Klieger, D. M., Borneman, M. J., & Kuncel, N. R. (2009). The predictive power of personal statements in admissions: A meta-analysis and cautionary tale. *College and University*, *84*, 83-86.
Naylor, J. C. & Shine, L. C. (1965). A table for determining the increase in mean criterion score obtained by using a selection device. *Journal of Industrial Psychology, 3*, 33-42.
Nguyen, N. T., Biderman, M. D., & McDaniel, M. A. (2005). Effects of response instructions on faking a situational judgment test. *International Journal of Selection and Assessment*, *13*, 250–260. doi:10.1111/j.1468-2389.2005.00322.x
Niessen, A. S. M. & Meijer, R. R. (2015). Wanneer heeft selectie in het hoger onderwijs zin? De stand van zaken anno 2015 [When does selection in higher education pay off? The situation in 2015]. *Tijdschrift voor Hoger Onderwijs, 33*, 4-19.
Niessen, A. S. M. & Meijer, R. R. (2016). Selection of medical students on the basis of non-academic skills: Is it worth the trouble? *Clinical Medicine, 16*, 339-342. doi:10.7861/clinmedicine.16-4-339
Niessen, A. S. M., & Meijer, R. R. (2017). On the use of broadened admission criteria in higher education. *Perspectives on Psychological Science*, *12*, 436-448. doi:10.1177/1745691616683050

Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2016). Predicting performance in higher education using proximal predictors. *PLoS ONE, 11*(4), e0153663. doi:10.1371/journal.pone.0153663

Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2017a). Applying organizational justice theory to admission into higher education: Admission from a student perspective. *International Journal of Selection and Assessment*, *25*, 72-84. doi:10.1111/ijsa.12161

Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2017b). Measuring noncognitive predictors in high-stakes contexts: The effect of self-presentation on self-report instruments used in admission to higher education. *Personality and Individual Differences, 106*, 183-189. doi:10.1016/j.paid.2016.11.014

Nikolaou, I., & Judge, T. A. (2007). Fairness reactions to personnel selection techniques in Greece: The role of core self-evaluations. *International Journal of Selection and Assessment*, *15*, 206-219. doi:10.1111/j.1468-2389.2007.00382.x

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41*, 673–690. doi:10.1007/s11135-006-9018-6

O'Neill, T. A., Goffin, R. D., & Gellatly, I. R. (2010). Test-taking motivation and personality test validity. *Journal of Personnel Psychology, 9*, 117–125. doi:10.1027/ 1866-5888/a000012.

O'Neill, T. A., Lewis, R. J., Law, S. J., Larson, N., Hancock, S., Radan, J., ... Carswell, J. J. (2017). Forced-choice pre-employment personality assessment: Construct validity and resistance to faking. *Personality and Individual Differences, 115*, 120-127. doi:10.1016/j.paid. 2016.03.075

Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology, 60*, 995-1027. doi:10.1111/j.1744-6570.2007.00099.x

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81,* 660–679. doi:10.1037/0021-9010.81.6.660.

Ooijevaar, J. (2010). Allochtonen en autochtonen in het hoger onderwijs [Ethnic minority- and ethnic mjority students in higher education]. *Sociaal econmische trends*, *4,* 37-42. CBS, the Netherlands. Retrieved from: https://www.cbs.nl/nl-nl/achtergrond/2010/10/allochtonen-en-autochtonen-in-het-hoger-onderwijs

Organisation for Economic Co-operation and Development. (2012). *Equity and quality in education: Supporting disadvantaged students and schools*. Paris, France: OECD Publishing. doi:10.1787/9789264130852-en

Organisation for Economic Co-operation and Development. (2014). *PISA 2012 results: Creative problem solving: Students' skills in tackling real-life problems* (Vol. *V*). Paris, France: OECD Publishing. doi:10.1787/9789264208070-en

Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology*, *89*, 187-207. doi:10.1037/0021-9010.89.2.187

Patterson, F., Ashworth, V., Zibarras, L., Coan, P., Kerrin, M., & O'Neill, P. (2012). Evaluations of situational judgement tests to assess non-academic attributes in selection. *Medical Education*, *46*, 850-868. doi:10.1111/j.1365-2923.2012.04336.x

Patterson, F., Baron, H., Carr, V., Plint, S., & Lane, P. (2009). Evaluation of three short-listing methodologies for selection into postgraduate training in general practice. *Medical Education, 43,* 50-57. doi:10.1111/j.1365-2923.2008.03238.x

Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousans, F., & Cleland, J. (2016). How effective are selection methods in medical education? A systematic review. *Medical Education, 50*, 36–60. doi:10.1111/medu.12817

Patterson, F., Zibarras, L., Carr, V., Irish, B., & Gregory, S. (2011). Evaluating candidate reactions to selection practices using organisational justice theory. *Medical Education*, *45*, 289-297. doi:10.1111/j.1365-2923.2010.03808.x

Pau, A., Jeevaratnam, K., Chen, Y. S., Fall, A. A., Khoo, C., & Nadarajah, V. D. (2013). The multiple mini-interview (MMI) for student selection in health professions training: A systematic review. *Medical Teacher*, *35*, 1027-1041. doi:10.3109/0142159X.2013.829912

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, & P. R. Shaver (Eds.), *Measures of personality and social psychological attitudes*. *Vol. 1.* (pp. 17–59). San Diego, CA: Academic Press.

Pauls, C. A., & Crost, N.W. (2004). Effects of faking on self-deception and impression management scales. *Personality and Individual Differences, 37,* 1137–1151. doi:10.1016/j.paid.2003.11.018.

Peers, I. S. & Johnston, M. (1994). Influence of learning context on the relationship between A-level attainment and final degree performance: A meta-analytic review. *British Journal of Educational Psychology, 64*, 1-18. doi:10.1111/j.2044-8279.1994.tb01081.x

Peeters, H., & Lievens, F. (2005). Situational judgment tests and their predictiveness of college students' success: The influence of faking. *Educational and Psychological Measurement, 65,* 70–89. doi:10.1177/0013164404268672

Peterson, M. H., Griffith, R. L., Isaacson, J. A., O'Connell, M. S., & Mangos, P. M. (2011). Applicant faking, social desirability, and the prediction of counterproductive work behaviors. *Human Performance, 24,* 270–290. doi:10.1080/08959285.2011.580808.

Ployhart, R. E., & Holtz, B. C. (2008). The diversity-validity dilemma: Strategies for reducing racio-ethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, *61*, 153-172. doi:10.1111/j.1744-6570.2008.00109.x

Ployhart, R. E., Schneider, B., & Schmitt, N. (2006). *Staffing organizations: Contemporary practice and theory* (3rd ed). Mahway, NJ: Lawrence Erlbaum Associates

Plummer, M. (2016a). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling [Computer software].

Plummer, M. (2016b). rjags: Bayesian Graphical Models using MCMC. R package version 4-6. Available from: http://CRAN.R-project.org/package=rjags

Pretz, J. E., & Kaufman, J. C. (2015). Do traditional admissions criteria reflect applicant creativity? *The Journal of Creative Behavior*, *51,* 240–251. doi:10.1002/jocb.120

Prevatt, F., Li, H., Welles, T., Festa-Dreher, D., Yelland, S., & Lee, J. (2011). The academic success inventory for college students: Scale development and practical implications for use with students. *Journal of College Admission*, *211*, 26–31.

Ramsay, L. J., Schmitt, N., Oswald, F. L., Kim, B. H., & Gillespie, M. A. (2006). The impact of situational context variables on responses to biodata and situational judgment inventory items. *Psychology Science*, *48*, 268-287.

Reardon, S. F., & Portilla, X. A. (2016). Recent trends in income, racial, and ethnic school readiness gaps at kindergarten entry. *AERA Open*, *2*, 1–18. doi:10.1177/2332858416657343

Reibnegger, G., Caluba, H. C., Ithaler, D., Manhal, S., Neges, H. M., & Smolle, J. (2010). Progress of medical students after open admission or admission based on knowledge tests. *Medical Education, 44,* 205–214. doi:10.1111/j.1365-2923.2009.03576.x

Reiter, H. I., Eva, K. W., Rosenfeld, J., & Norman, G. R. (2007). Multiple mini-interviews predict clerkship and licensing examination performance. *Medical Education*, *41*, 378-384. doi:10.1111/j.1365-2929.2007.02709.x

Resing, W. C. M., & Drenth, P. J. D. (2007). *Intelligentie: Weten en meten* [Intelligence: Measuring and knowing]. Amsterdam, the Netherlands: Uitgeverij Nieuwezijds.

Revelle, W. (2015) *Psych: Procedures for personality and psychological research.* R package version 1.5.8. Retrieved from: http://CRAN.R-project.org/package=psych

Richardson, M., Abrahams, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis*. Psychological Bulletin, 138,* 353–387. doi:10.1037/a0026838.

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, *130*, 261-288. doi:10.1037/0033-2909.130.2.261

Robertson, I. T., & Kandola, R. S. (1982). Work sample tests: Validity, adverse impact and applicant reaction. *Journal of Occupational Psychology*, *55*, 171-183. doi:10.1111/j.2044-8325.1982.tb00091.x

Roberts, B.W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin, 132,* 1–25. doi:10.1037/0033-2909.132.1.1.

Roe, R. A. (1990). Personeelsselectie: modellen en instrumenten [Personnel selection: models and instruments]. In: P.J.D. Drenth, Hk. Thierry en Ch. J.

de Wolff (Eds.), *Nieuw Handboek Arbeids- en Organisatiepsychologie.* Deventer, the Netherlands: van Loghum Slaterus
Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology, 83,* 634–644. doi:02I-90IO/98/J3.00
Roth, P. L., Bobko, P., McFarland, L., & Buster, M. (2008). Work sample tests in personnel selection: A meta-analysis of black-white differences in overall and exercise scores. *Personnel Psychology, 61,* 637–661. doi:10.1111/j.1744-6570.2008.00125.x
Roth, P. L., Buster, M. A., & Barnes-Farrell, J. (2010). Work sample exams and gender adverse impact potential: The influence of self-concept, social skills, and written skills. *International Journal of Selection and Assessment*, *18*, 117-130. doi:10.1111/j.1468-2389.2010.00494.x
Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support? *Human Resource Management Review*, *16*, 155–180. doi:10.1016/j.hrmr.2006.03.004
Ryan, A. M., & Boyce, A. S. (2006). What do we know and where do we go? Practical directions for faking research. In R. L. Griffith, & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 357–371). Greenwich, CT: Information Age.
Ryan, A.M., Inceoglu, I., Bartram, D., Golubovich, Y., Grand, J., Reeder,M., ... Yao, X. (2015). Trends in testing: Highlights of a global survey. In I. Nikolaou, & J. K. Oostrom (Eds.), *Employee recruitment, selection, and assessment: Contemporary issues for theory and practice* (pp. 136–153). New York, NY: Psychology Press/Taylor & Francis.
Ryan, A. M., McFarland, L. A., Baron, H., & Page, R. (1999). An international look at selection practices: Nation and culture as Explanations for variability in practice. *Personnel Psychology, 52,* 359-391. doi:10.1111/j.1744-6570.1999.tb00165.x
Ryan, A. M., & Ployhart, R. E. (2000). Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management*, *26*, 565-606. doi:10.1177/014920630002600308
Ryan, A. M., Sacco, J. M., McFarland, L. A., & Kriska, S. D. (2000). Applicant self-selection: Correlates of withdrawal from a multiple hurdle process. *Journal of Applied Psychology*, *85*, 163-179. doi:10.1037/0021-9010.85.2.163
Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2008). High stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *American Psychologist*, *63*, 215-227. doi:10.1037/0003-066X.63.4.215
Sackett, P. R., & Ellingson, J. E. (1997). The effects of forming multi-predictor composites on group differences and adverse impact. *Personnel Psychology*, *50*, 707–721. doi:10.1111/j.1744-6570.1997.tb00711.x
Sackett, P. R., Kuncel, N. R., Arneson, J. J., Cooper, S. R., & Waters, S. D. (2009). Does socioeconomic status explain the relationship between admissions tests

and post-secondary academic performance? *Psychological Bulletin, 135*, 1–22. doi:10.1037/a0013978

Sackett, P. R., Laczo, R. M., & Lippe, Z. P. (2003). Differential prediction and the use of multiple predictors: The omitted variables problem. *Journal of Applied Psychology*, *88*, 1046-1056. doi:10.1037/0021-9010.88.6.1046

Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education. *American Psychologist*, *56*, 302–318. doi:I0.IO37/AJO03-O66X.56.4.302

Sackett, P. R., Walmsley, P. T., Koch, A. J., Beatty, A. S., & Kuncel, N. R. (2016). Predictor content matters for knowledge testing: Evidence supporting content validation. *Human Performance*, *29*, 54-71. doi:10.1080/08959285.2015.1120307

Sanchez, R. J., Truxillo, D. M., & Bauer, T. N. (2000). Development and examination of an expectancy-based measure of test-taking motivation. *Journal of Applied Psychology*, *85*, 739-750. doi:10.1037/0021-9010.85.5.739

Schmidt, F. L. & Hunter, J. E. (1998) The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262–274. doi:10.1037/0033-2909.124.2.262

Schmit, M. J., & Ryan, A. M. (1993). The big-5 in personnel-selection: Factor structure in applicant and non-applicant populations. *Journal of Applied Psychology, 78*, 966–974. doi:10.1037/0021-9010.78.6.966

Schmitt, N. (2012). Development of rationale and measures of noncognitive college student potential. *Educational Psychologist, 47,* 18-29. doi:10.1080/00461520.2011.610680

Schmitt, N., Clause, C. S., & Pulakos, E. D. (1996). Subgroup differences associated with different measures of some common job relevant constructs. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (pp. 115–140). New York, NY: Wiley.

Schmitt, N., Keeney, J., Oswald, F. L., Pleskac, T. J., Billington, A. Q., Sinha, R., & Zorzie, M. (2009). Prediction of 4-year college student performance using cognitive and noncognitive predictors and the impact on demographic status of admitted students. *Journal of Applied Psychology*, *94*, 1479– 1497. doi:10.1037/a0016810

Schmitt, N., & Mills, A. E. (2001). Traditional tests and job simulations: Minority and majority performance and test validities. *Journal of Applied Psychology*, *86*, 451-458. doi:10.1037/0021-9010.86.3.451

Schreurs, B., Derous, E., Proost, K., Notelaers, G., & De Witte, K. (2008). Applicant selection expectations: Validating a multidimensional measure in the military. *International Journal of Selection and Assessment*, *16*, 170-176. doi:10.1111/j.1468-2389.2008.00421.x

Schripsema, N. R., van Trigt, A. M., Borleffs, J. C. C., & Cohen-Schotanus, J. (2014). Selection and study performance: comparing three admission processes within one medical school. *Medical Education, 48,* 1201-1210. doi:10.1111/medu.12537

Schwager, I. T. L., Hülsheger, U. R., Bridgeman, B., & Lang, J. W. B. (2015). Graduate student selection: Graduate record examination, socioeconomic status, and

undergraduate grade point average as predictors of study success in a western European University. *International Journal of Selection and Assessment, 23*, 71-79. doi:10.1111/ijsa.12096

Schwartz, B. (2015, September 25). Do college admissions by lottery. *The New York Times*. Retrieved from: https://www.nytimes.com/roomfordebate/2015/03/31/how-to-improve-the-college-admissions-process/do-college-admissions-by-lottery

Shen, W., Sackett, P. R., Kuncel, N. R., Beatty, A. S., Rigdon, J L., & Kiger, T. B. (2012). All validities are not created equal: Determinants of variation in sat validity across schools. *Applied Measurement in Education, 25*, 197-219. doi:10.1080/08957347.2012.687615

Shultz, M. M., & Zedeck, S. (2012). Admission to law school: New measures. *Educational Psychologist*, *47*, 51-65. doi:10.1080/00461520.2011.610679

Sijtsma, K. & Evers, A. (2011). Validiteit als tweezijdig problem [Validity as a two-sided problem]. *De Psycholoog, 46*, 40-46.

Sinha, R., Oswald, F., Imus, A., & Schmitt, N. (2011). Criterion-focused approach to reducing adverse impact in college admissions. *Applied Measurement in Education*, *24*, 137–161. doi:10.1080/08957347.2011.554605

Smith, M. (1948). Cautions concerning the use of the Taylor-Russell tables in employee selection. *Journal of Applied Psychology*, *32*, 595-600. doi:10.1037/h0058507

Smither, J. W., Reilly, R. R., Millsap, R. E., & Pearlman, K. (1993). Applicant reactions to selection procedures. *Personnel Psychology*, *46*, 49-76. doi:10.1111/j.1744-6570.1993.tb00867.x

Smits, N. (2003). *Academic specialization choices and academic achievement: Prediction and incomplete data* (doctoral dissertation). Amsterdam: University of Amsterdam.

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2011). Constructing fake-resistant personality tests using item response theory: High-stakes personality testing with multidimensional pairwise preferences. In M. Ziegler, C. MacCann, R. D. Roberts, M. Ziegler, C. MacCann, R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 214-239). New York, NY, US: Oxford University Press.

Stark, S., Chernyshenko, O. S., Drasgow, F., Nye, C. D., White, L. A., Heffner, T., & Farmer, W. L. (2014). From ABLE to TAPAS: A new generation of personality tests to support military selection and classification decisions. *Military Psychology, 26*, 153–164. doi:10.1037/mil0000044.

Statistics Netherlands. (2014). *WO; studievoortgang, vooropleiding, studierichting, herkomstgroepering* [Research universities, study progress, previous education, academic program, and background] [Data file]. Retrieved from: http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=71199ned&D1=0-21,30-53&D2=0&D3=0&D4=0&D5=a&D6=0&D7=a&VW=T

Steele-Johnson, D., & Leas, K. (2013). Importance of race, gender, and personality in predicting academic performance. *Journal of Applied Social Psychology*, *43*, 1736-1744. doi:10.1111/jasp.12129

Stegers-Jager, K. M., Steyerberg, E. W., Lucieer, S. M., & Themmen, A. N. (2015). Ethnic and social disparities in performance on medical school selection criteria. *Medical Education*, *49*, 124-133. doi:10.1111/medu.12536

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87,* 45-51. doi:10.1037/0033-2909.87.2.245

Steiner, D. D., & Gilliland, S. W. (1996). Fairness reactions to personnel selection techniques in France and the United States. *Journal of Applied Psychology*, *81*, 134-141. doi:10.1037/0021-9010.81.2.134

Steiner, D. D., & Gilliland, S. W. (2001). Procedural justice in personnel selection: International and crosscultural perspectives. *International Journal of Selection and Assessment*, *9*, 124-137. doi:10.1111/1468-2389.00169

Stemig, M. S., Sackett, P. R., & Lievens, F. (2015). Effects of organizationally endorsed coaching on performance and validity of situational judgment tests. *International Journal of Selection and Assessment*, *23*, 174–181. doi:10.1111/ijsa.12105

Stemler, S. E. (2012). What should university admissions tests predict? *Educational Psychologist*, *47*, 5-17. doi:10.1080/00461520.2011.611444

Stemler, S. E. (2017). College admissions, the MIA model, and MOOCs: Commentary on Niessen and Meijer (2017). *Perspectives on Psychological Science, 12*, 449-451. doi:10.1177/1745691617690879

Sternberg, R. J. (2003). Our research program validating the triarchic theory of successful intelligence: Reply to Gottfredson. *Intelligence*, *31*, 399–413. doi:10.1016/S0160-2896(02)00143-5

Sternberg, R. J. (2010). College admissions for the 21st century. Cambridge, MA: Harvard University Press.

Sternberg, R. J., Bonney, C. R., Gabora, L., Jarvin, L., Karelitz, T. M., & Coffin, L. (2010). Broadening the spectrum of undergraduate admissions: The Kaleidoscope Project. *College and University*, *86*, 2–17.

Sternberg, R. J., Bonney, C. R., Gabora, L., & Merrifield, M. (2012). WICS: A model for college and university admissions. *Educational Psychologist*, *47*, 30–34. doi:10.1080/00461520.2011.638882

Sternberg, R. J. & The Rainbow Project Collaborators. (2006). The Rainbow Project: Enhancing the SAT through assessments of analytical, practical, and creative skills. *Intelligence*, *34*, 321–350. doi:10.1016/j.intell.2006.01.002

Stone, P. (2008a). On fair lotteries. *Social Theory and Practice, 34,* 573-590. doi:10.5840/soctheorpract200834431

Stone, P. (2008b). What can lotteries do for education? *Theory and Research in Education, 6,* 267–282. doi:10.1177/1477878508095583

Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: discussion and tables. *Journal of Applied Psychology*, *23*, 565-578. doi:10.1037/h0057079

Tett, R. P., Freund, K. A., Christiansen, N. D., Fox, K. E., & Coaster, J. (2012). Faking on self-report emotional intelligence and personality tests: Effects of

faking opportunity, cognitive ability, and job type. *Personality and Individual Differences, 52*, 195–201. doi:10.1016/j.paid.2011.10.017.
Thorndike, E. L. (1906). An empirical study of college entrance examinations. *Science, 23,* 839-845. doi:10.1126/science.23.596.839
Thornton, G. & Kedharnath, U. (2013). Work sample tests. In: K. F. Geisinger, B. A. Bracken, J. F. Carlson, J. C. Hansen, N. R. Kuncel, S. P. Reise, et al., (Eds.). *APA handbook of testing and assessment in psychology, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology* (pp. 533-550). Washington, DC: American Psychological association.
Thorsteinson, T. J. & Ryan, A.M. (1997). The effect of selection ratio on the perceptions of the fairness of a selection test battery. *International Journal of Selection and Assessment, 5,* 159-168. doi:10.1111/1468-2389.00056
Thurstone, L. L. (1919). Mental tests for college entrance. *Journal of Educational Psychology*, *10*, 129-142. doi:10.1037/h0071432
Topping, J. D., & O'Gorman, J. G. (1997). Effects of faking set on validity of the NEO-FFI. *Personality and Individual Differences*, *23*, 117–124. doi:10.1016/S01918869(97)00006-8
Truijens, A. (2014, September 21). Loting voor studie is een rotsysteem, maar misschien is er geen beter [Lottery admission is awful, but perhaps there is no better alternative]. *De Volkskrant.* Retrieved from: https://www.volkskrant.nl/dossier-aleid-truijens/loting-voor-studie-is-een-rotsysteem-maar-misschien-is-er-geen-beter~a3751894/
Truxillo, D. M., Bodner, T. E., Bertolino, M., Bauer, T. N., & Yonce, C. A. (2009). Effects of explanations on applicant reactions: A meta-analytic review. *International Journal of Selection and Assessment*, *17*, 346-361. doi:10.1111/j.1468-2389.2009.00478.x
Truxillo, D. M., Steiner, D. D., & Gilliland, S. W. (2004). The importance of organizational justice in personnel selection: defining when selection fairness really matters. *International Journal of Selection and Assessment*, *12*, 39-53. doi:10.1111/j.0965-075X.2004.00262.x
Urlings-Strop, L. C., Stegers-Jager, K. M., Stijnen, T., & Themmen, A. N. (2013). Academic and non-academic selection criteria in predicting medical school performance. *Medical Teacher*, *35*, 497-502. doi:10.3109/0142159X.2013.774333
Urlings-Strop, L. C., Stijnen, T., Themmen, A. N., & Splinter, T. W. (2009). Selection of medical students: A controlled experiment. *Medical Education*, *43*, 175-183. doi:10.1111/j.1365-2923.2008.03267.x
Valli, R. & Johnson, P. (2013). Entrance examinations as gatekeepers. *Scandinavian Journal of Educational Research, 51*, 493–510. doi:10.1080/00313830701576631
van Bebber, J., Lem, J., & van Zoelen, L. (2010). *Q1000 Capaciteiten Hoog* [Q1000 High Capacities]. Woerden, the Netherlands: Meurs HRM
van den Broek, A., Kerstens, J. & Woutersen, M. (2003). Lot in eigen hand: Evaluatie van de experimenten met decentrale toelating in het hoger onderwijs [The lot in their own hands: Evaluation of selective admisison experiments in higher education]. Nijmegen, the Netherlands: IOWO

van den Broek, A., Nooij, J., van Essen, M., & Duysak, S. (2017). *Selectie & plaatsing bij numerusfixusopleidingen* [Selection and placement for numerus fixus programs]. Nijmegen, the Netherlands: ResearchNed. Retrieved from: https://www.rijksoverheid.nl/documenten/rapporten/2017/06/01/selectie-en-plaatsing-bij-numerusfixusopleidingen

van der Flier, H. (1992). *Hebben wij eigenschappen nodig? Signs en samples in het psychologisch selectie-onderzoek* [Do we need characteristics? Signs and samples in psychological selection research]. Amsterdam, The Netherlands: VU University

van der Heijden, P. G. M., Hessen, D.J., & Wubbels, T. (2012). Studiesucces of -falen van eerstejaarsstudenten voorspellen: een nieuwe aanpak [Predicting academic succes or failure: a new approach]. *Tijdschrift voor Hoger Onderwijs, 4*, 233-244.

van der Maas, H. & Visser, K. (2017, June 8). Wet selectie studenten is niet uitvoerbaar [Admission law is not feasible]. *De Volkskrant.* Retrieved from: https://www.volkskrant.nl/opinie/wet-selectie-studenten-is-niet-uitvoerbaar~a4499574/

Vasilopoulos, N. L., Cucina, J. M., Dyomina, N. V., Morewitz, C. L., & Reilly, R. R. (2006). Forced-choice personality tests: A measure of personality and cognitive ability? *Human Performance*, *19*, 175–199. doi:10.1207/s15327043hup1903_1

Vickers, J. M. (2000). Justice and truth in grades and their averages. *Research in Higher Education*, *41*, 141-64. doi:10.1023/A:1007069620329

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*, 1–48. doi:10.18637/jss.v036.i03

Vihavainen, A., Luukkainen, M., & Kurhila, J. (2013, October). *MOOC as semester-long entrance exam*. Paper presented at the 14th Annual ACM SIGITE Conference on Information Technology Education, Orlando, Florida, United States.

Visser, K. (2017, August 23). Kantel de selectie in het hoger onderwijs [Topple selective admission in higher education]. *ScienceGuide.* Retrieved from: http://www.scienceguide.nl/201708/kantel-de-selectie-in-het-hoger-onderwijs.aspx

Visser, K., van der Maas, H., Engels-Freeke, M., & Vorst, H. (2012). Het effect op studiesucces van decentrale selectie middels proefstuderen aan de poort [The effect on study success of student selection through trial-studying]. *Tijdschrift voor Hoger Onderwijs, 30*, 161–173.

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59*, 197–210. doi:10.1177/00131649921969802.

von Stumm, S. & Ackerman, P. L. (2013). Investment and intellect: A review and meta-analysis. *Psychological Bulletin*, *139*, 841-869. doi:10.1037/a0030746

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*, 779-804. doi:10.3758/BF03194105

Wagerman, S. A., & Funder, D. C. (2007). Acquaintance reports of personality and academic achievement: A case for conscientiousness. *Journal of Research in Personality*, *41*, 221–229. doi:10.1016/j.jrp.2006.03.001

Warps, J., Nooij, N., Muskens, M., Kurver, B., & van den Broek, A. (2017). *De studiekeuzecheck* [The study choice check]. Nijmegen, the Netherlands: ResearchNed. Retrieved from: https://www.rijksoverheid.nl/documenten/rapporten/2017/01/20/de-studiekeuzecheck-landelijk-onderzoek-naar-uitvoering-en-opbrengsten-van-de-studiekeuzecheck

Weigold, I. K., Weigold, A., Kim, S., Drakeford, N. M., & Dykema, S. A. (2016). Assessment of the psychometric properties of the Revised Academic Hardiness Scale in college student samples. *Psychological Assessment*, *28*, 1207–1219. doi:10.1037/pas0000255

Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology*, *52*, 372-376. doi:10.1037/h0026244

Westrick, P. A., Le, H., Robbins, S. B., Radunzel, J. R., & Schmidt, F. L. (2015). College performance and retention: A meta-analysis of the predictive validities of ACT scores, high school grades, and SES. *Educational Assessment*, *20*, 23-45. doi:10.1080/10627197.2015.997614

Wet kwaliteit in verscheidenheid [Quality in diversity Law] (2013). *Memorie van Toelichting [Note of explanation].* Retrieved from: https://www.eerstekamer.nl/behandeling/20130118/memorie_van_toelichting_2/document3/f=/vj6ifr0xnkyo.pdf

Wetzel, E., Bohnke, J. R., & Brown, A. (2016). Response biases. In F. T. L. Leong, D. Bartram, F. Cheung, K. F. Geisinger, & D. Iliescu (Eds.), *The ITC international handbook of testing and assessment* (pp. 349–363). Oxford, UK: Oxford University Press.

Wilbrink, B. (1973). Tegenspraken typeren het selektie-voorstel De Groot [Selection propsal by de Groot meets criticism]. *OTO Hoger Onderwijs Cahiers, 13*, 166-176.

Wolfe, R. N., & Johnson, S. D. (1995). Personality as a predictor of college performance. *Educational and Psychological Measurement*, *55*, 177–185. doi:10.1177/0013164495055002002

Wolming, S. (1999). Validity issues in higher education selection: A Swedish example. *Studies in Educational Evaluation 25,* 335 351. doi:10.1016/S0191-491X(99)00034-6

Wouters, A. (2017). *Effects of medical school selection on the motivation of the student population and the applicant pool* (doctoral dissertation). Amsterdam, the Netherlands, VU University

Wouters, A., Croiset, G., Schripsema, N. R., Cohen-Schotanus, J., Spaai, G. W. G., Hulsman, R. L., & Kusurkar, R. A. (2017). Students' approaches to medical school choice: relationship with students' characteristics and motivation. *International Journal of Medical Education*, *8*, 217–226. doi:10.5116/ijme.5921.5090

Young, J. W. (2007). Predicting college grades: The value of achievement goals in supplementing ability measures. *Assessment in Education: Principles, Policy & Practice*, *14*, 233–249. doi:10.1080/09695940701479709

Zickar, M. J., & Robie, C. (1999). Modeling faking good on personality items: An item-level analysis. *Journal of Applied Psychology, 84*, 551–563. doi:10.1037/0021- 9010.84.4.551.

Ziegler, M. (2011). Applicant faking: A look into the black box. *The industrial and organizational psychologist, 49*, 29–36.

Ziegler, M., Danay, E., Schölmerich, F., & Bühner, M. (2010). Predicting academic success with the big 5 rated from different points of view: Self-rated, other rated and faked. *European Journal of Personality, 24*, 341–355. doi:10.1002/per.753.

Zwick, R. (2007). College admissions in twenty-first-century America: The role of grades, tests, and games of chance. *Harvard Educational Review, 77*, 419-429. doi:10.17763/haer.77.4.u67n84589527t80v

Zwick, R. (2012). The role of admissions test scores, socioeconomic status, and high school grades in predicting college achievement. *Pensamiento Educativo, Revista de Investigación Educacional Latinoamericana*, *49*, 23–30.

Zwick, R. (2013). *Disentangling the role of high school grades, SAT scores, and SES in predicting college achievement.* ETS Research Report Series (Report no. ETS RR–13-09). Princeton, NJ: Educational Testing Service

Zwick, R. (2017). *Who gets in?: Strategies for fair and effective college admissions*. Cambridge, MA: Harvard University Press

Zwick, R. & Himmelfarb, I. (2011). The effect of high school socioeconomic status on the predictive validity of SAT scores and high school grade-point average. *Journal of Educational Measurement, 48*, 101–121. doi:10.1111/j.1745-3984.2011.00136.x